

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- 00360934-073599
- 669220 "HE60934"
- (i) APPLICANT: Covacci, Antonello
Bugnoli, Massimo
Telford, John
Macchia, Giovanni
Rappuoli, Rino
 - (ii) TITLE OF INVENTION: Helicobacter Pylori Proteins Useful
for Vaccines and Diagnostics
 - (iii) NUMBER OF SEQUENCES: 7
 - (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Chiron Corporation
 - (B) STREET: 4560 Horton Street
 - (C) CITY: Emeryville
 - (D) STATE: California
 - (E) COUNTRY: USA
 - (F) ZIP: 94608-2916
 - (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
 - (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER: 08/471,491
 - (B) FILING DATE: 06-JUNE-1995
 - (C) CLASSIFICATION:
 - (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: McClung, Barbara G.
 - (B) REGISTRATION NUMBER: 33,113
 - (C) REFERENCE/DOCKET NUMBER: 0316.003
 - (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: (510) 601-2708
 - (B) TELEFAX: (510) 655-3542

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 27 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GCAAGCTTAT CGATGTCGAC TCGAGCT

27

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3960 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

AAAAAGAAAG GAAGAAAATG GAAATACAAC AAACACACCG CAAAATCAAT CGCCCTCTGG	60
TTTCTCTCGC TTTAGTAGGA GCATTAGTCA GCATCACACC GCAACAAAGT CATGCCGCCT	120
TTTTCACAAC CGTGATCATT CCAGCCATTG TTGGGGGTAT CGCTACAGGC ACCGCTGTAG	180
GAAEGGTCTC AGGGCTTCTT AGCTGGGGGC TCAAACAAGC CGAAGAAGCC AATAAAACCC	240
CAGATAAACC CGATAAAGTT TGGCGCATTC AAGCAGGAAA AGGCTTTAAT GAATTCCTTA	300
ACAAGGAATA CGACTTATAC AGATCCCTTT TATCCAGTAA GATTGATGGA GGTGTTGGATT	360
GGGGGAATGC CGCTAGGCAT TATTGGGTCA AAGGCGGGCA ACAGAATAAG CTTGAAGTGG	420
ATATGAAAGA CGCTGTAGGG ACTTATACCT TATCAGGGCT TAGAACTTT ACTGGTGGGG	480
ATTTAGATGT CAATATGCAA AAAGCCACTT TACGCTTGGG CCAATTCAAT GGCAATTCTT	540
TTACAAGCTA TAAGGATAGT GCTGATCGCA CCACGAGAGT GGATTTCAAC GCTAAAATA	600
TCTCAATTGA TAATTTTGTA GAAATCAACA ATCGTGTGGG TTCTGGAGCC GGGAGGAAAG	660
CCAGCTCTAC GGTTTGTGACT TTGCAAGCTT CAGAAGGGAT CACTAGCGAT AAAAACGCTG	720
AAATTTCTCT TTATGATGGT GCCACGCTCA ATTTGGCTTC AAGCAGCGTT AAATTAATGG	780
GTAATGTGTG GATGGGCCGT TTGCAATACG TGGGAGCGTA TTTGGCCCCT TCATACAGCA	840
CGATAAACAC TTCAAAAGTA ACAGGGGAAG TGAATTTTAA CCACCTCACT GTTGCGGATA	900

AAAACGCCGC	TCAAGCGGGC	ATTATCGCTA	ATAAAAAGAC	TAATATTGGC	ACACTGGATT	960
TGTGGCAAAG	CGCCGGGTTA	AACATTATCG	CTCCTCCAGA	AGGTGGCTAT	AAGGATAAAC	1020
CCAATAATAC	CCCTTCTCAA	AGTGGTGCTA	AAAACGACAA	AAATGAAAGC	GCTAAAAACG	1080
ACAAACAAGA	GAGCAGTCAA	AATAATAGTA	ACACTCAGGT	CATTAACCCA	CCCAATAGTG	1140
CGCAAAAAAC	AGAAGTTCAA	CCCACGCAAG	TCATTGATGG	GCCTTTTGCG	GGCGGCAAAG	1200
ACACGGTTGT	CAATATCAAC	CGCATCAACA	CTAACGCTGA	TGGCACGATT	AGAGTGGGAG	1260
GGTTTAAAGC	TTCTCTTACC	ACCAATGCGG	CTCATTTGCA	TATCGGCAA	GGCGGTGTCA	1320
ATCTGTCCAA	TCAAGCGAGC	GGGCGCTCTC	TTATAGTGGA	AAATCTAACT	GGGAATATCA	1380
CCGTTGATGG	GCCTTTAAGA	GTGAATAATC	AAGTGGGTGG	CTATGCTTTG	GCAGGATCAA	1440
GCGGGAATTT	TGAGTTTAAG	GCTGGTACGG	ATACCAAAAA	CGGCACAGCC	ACTTTTAATA	1500
ACGATATTAG	TCTGGGAAGA	TTTGTGAATT	TAAAGGTGGA	TGCTCATACA	GCTAATTTTA	1560
AAGGTATTGA	TACGGGTAAT	GGTGGTTTCA	ACACCTTAGA	TTTLAGTGGC	GTTACAGACA	1620
AAGTCAATAT	CAACAAGCTC	ATTACGGCTT	CCACTAATGT	GGCCGTTAAA	AACTTCAACA	1680
TTAATGAATT	GATTGTTAAA	ACCAATGGGA	TAAGTGTGGG	GGAATATACT	CATTTTAGCG	1740
AAGATATAGG	CAGTCAATCG	CGCATCAATA	CCGTGCGTTT	GGAAACTGGC	ACTAGGTCAC	1800
TTTCTCTGG	GGGTGTTAAA	TTTAAAGGTG	GCGAAAAATT	GGTTATAGAT	GAGTTTTACT	1860
ATAGCCCTTG	GAATTATTTT	GACGCTAGAA	ATATTAAAAA	TGTTGAAATC	ACCAATAAAC	1920
TTGCTTTTGG	ACCTCAAGGA	AGTCCTTGGG	GCACATCAAA	ACTTATGTTC	AATAATCTAA	1980
CCCTAGGTCA	AAATGCGGTC	ATGGATTATA	GCCAATTTTC	AAATTTAACC	ATTCAAGGGG	2040
ATTTCAATCA	CAATCAAGGC	ACTATCAACT	ATCTGGTCCG	AGGTGGGAAA	GTGGCAACCT	2100
TAAGCGTAGG	CAATGCAGCA	GCTATGATGT	TTAATAATGA	TATAGACAGC	GCGACCGGAT	2160
TTTACAAACC	GCTCATCAAG	ATTAACAGCG	CTCAAGATCT	CATTAAAAAT	ACAGAACATG	2220
TTTTATTGAA	AGCGAAAATC	ATTGGTTATG	GTAATGTTTC	TACAGGTACC	AATGGCATT	2280
GTAATGTAA	TCTAGAAGAG	CAATTCAAAG	AGCGCCTAGC	CCTTTATAAC	AACAATAACC	2340
GCATGGATAC	TTGTGTGGTG	CGAAATACTG	ATGACATTAA	AGCATGCGGT	ATGGCTATCG	2400
GCGATCAAAG	CATGGTGAAC	AACCCTGACA	ATTACAAGTA	TCTTATCGGT	AAGGCATGGA	2460

AAAATATAGG	GATCAGCAAA	ACAGCTAATG	GCTCTAAAAT	TTCGGTGTAT	TATTTAGGCA	2520
ATTCTACGCC	TACTGAGAAT	GGTGGCAATA	CCACAAATTT	ACCCACAAAC	ACCACTAGCA	2580
ATGCACGTTT	TGCCAACAAAC	GCCCTTGCAC	AAAACGCTCC	TTTCGCTCAA	CCTAGTGCTA	2640
CTCCTAATTT	AGTCGCTATC	AATCAGCATG	ATTTTGGCAC	TATTGAAAGC	GTGTTTGAAT	2700
TGGCTAACCG	CTCTAAAGAT	ATTGACACGC	TTTATGCTAA	CTCAGGCGCT	CAAGGCAGGG	2760
ATCTCTTACA	AACCTTATTG	ATTGATAGCC	ATGATGCGGG	TTATGCCAGA	AAAATGATTG	2820
ATGCTACAAG	CGCTAATGAA	ATCACCAAGC	AATTGAATAC	GGCCACTACC	ACTTTAAACA	2880
ACATAGCCAG	TTTAGAGCAT	AAAACCAGCG	GCTTACAAAC	TTTGAGCTTG	AGTAATGCGA	2940
TGATTTTAAA	TTCTCGTTTA	GTCAATCTCT	CCAGGAGACA	CACCAACCAT	ATTGACTCGT	3000
TCGCCAAACG	CTTACAAGCT	TTAAAAGACC	AAAAATTTCG	TTCTTTAGAA	AGCGCGGCAG	3060
AAGTGTGTGA	TCAATTTGCC	CCTAAATATG	AAAAACCTAC	CAATGTTTGG	GCTAACGCTA	3120
TTGGGGGAAC	GAGCTTGAAT	AATGGCTCTA	ACGCTTCATT	GTATGGCACA	AGCGCGGGCG	3180
TAGACGCTTA	CCTTAACGGG	CAAGTGGAAG	CCATTGTGGG	CGGTTTTGGA	AGCTATGGTT	3240
ATAGCTCTTT	TAATAATCGT	GCGAACTCCC	TTAACTCTGG	GGCCAATAAC	ACTAATTTTG	3300
GCGTGTATAG	CCGTATTTTT	GCCAACCAGC	ATGAATTGTA	CTTTGAAGCT	CAAGGGGCAC	3360
TAGGGAGCGA	TCAATCAAGC	TTGAATTTC	AAAGCGCTCT	ATTACAAGAT	TTGAATCAAA	3420
GCTATCATT	CTTAGCCTAT	AGCGCTGCAA	CAAGAGCGAG	CTATGGTTAT	GACTTCGCGT	3480
TTTTTAGGAA	CGCTTTAGTG	TTAAAACCAA	GCGTGGGTGT	GAGCTATAAC	CATTTAGGTT	3540
CAACCAACTT	TAAAAGCAAC	AGCACCAATC	AAGTGGCTTT	GAAAAATGGC	TCTAGCAGTC	3600
AGCATTTATT	CAACGCTAGC	GCTAATGTGG	AAGCGCGCTA	TTATTATGGG	GACACTTCAT	3660
ACTTCTACAT	GAATGCTGGA	GTTTTACAAG	AGTTCGCTCA	TGTTGGCTCT	AATAACGCCG	3720
CGTCTTTAAA	CACCTTTAAA	GTGAATGCCG	CTCGCAACCC	TTTAAATACC	CATGCCAGAG	3780
TGATGATGGG	TGGGGAATTA	AAATTAGCTA	AAGAAGTGTT	TTTGAATTTG	GGCGTTGTTT	3840
ATTTGCACAA	TTTGATTTC	AATATAGGCC	ATTTGCTTC	CAATTTAGGA	ATGAGGTATA	3900
GTTTCTAAAT	ACCGCTCTTA	AACCCATGCT	CAAAGCATGG	GTTTGAAATC	TTACAAAACA	3960

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1296 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Met Glu Ile Gln Gln Thr His Arg Lys Ile Asn Arg Pro Leu Val Ser
 1 5 10 15
 Leu Ala Leu Val Gly Ala Leu Val Ser Ile Thr Pro Gln Gln Ser His
 20 25 30
 Ala Ala Phe Phe Thr Thr Val Ile Ile Pro Ala Ile Val Gly Gly Ile
 35 40 45
 Ala Thr Gly Thr Ala Val Gly Thr Val Ser Gly Leu Leu Ser Trp Gly
 50 55 60
 Leu Lys Gln Ala Glu Glu Ala Asn Lys Thr Pro Asp Lys Pro Asp Lys
 65 70 75 80
 Val Trp Arg Ile Gln Ala Gly Lys Gly Phe Asn Glu Phe Pro Asn Lys
 85 90 95
 Glu Tyr Asp Leu Tyr Arg Ser Leu Leu Ser Ser Lys Ile Asp Gly Gly
 100 105 110
 Trp Asp Trp Gly Asn Ala Ala Arg His Tyr Trp Val Lys Gly Gly Gln
 115 120 125
 Gln Asn Lys Leu Glu Val Asp Met Lys Asp Ala Val Gly Thr Tyr Thr
 130 135 140
 Leu Ser Gly Leu Arg Asn Phe Thr Gly Gly Asp Leu Asp Val Asn Met
 145 150 155 160
 Gln Lys Ala Thr Leu Arg Leu Gly Gln Phe Asn Gly Asn Ser Phe Thr
 165 170 175
 Ser Tyr Lys Asp Ser Ala Asp Arg Thr Thr Arg Val Asp Phe Asn Ala
 180 185 190
 Lys Asn Ile Ser Ile Asp Asn Phe Val Glu Ile Asn Asn Arg Val Gly
 195 200 205
 Ser Gly Ala Gly Arg Lys Ala Ser Ser Thr Val Leu Thr Leu Gln Ala

0060934"072699

210 215 220
 Ser Glu Gly Ile Thr Ser Asp Lys Asn Ala Glu Ile Ser Leu Tyr Asp
 225 230 235 240
 Gly Ala Thr Leu Asn Leu Ala Ser Ser Ser Val Lys Leu Met Gly Asn
 245 250 255
 Val Trp Met Gly Arg Leu Gln Tyr Val Gly Ala Tyr Leu Ala Pro Ser
 260 265 270
 Tyr Ser Thr Ile Asn Thr Ser Lys Val Thr Gly Glu Val Asn Phe Asn
 275 280 285
 His Leu Thr Val Gly Asp Lys Asn Ala Ala Gln Ala Gly Ile Ile Ala
 290 295 300
 Asn Lys Lys Thr Asn Ile Gly Thr Leu Asp Leu Trp Gln Ser Ala Gly
 305 310 315 320
 Leu Asn Ile Ile Ala Pro Pro Glu Gly Gly Tyr Lys Asp Lys Pro Asn
 325 330 335
 Asn Thr Pro Ser Gln Ser Gly Ala Lys Asn Asp Lys Asn Glu Ser Ala
 340 345 350
 Lys Asn Asp Lys Gln Glu Ser Ser Gln Asn Asn Ser Asn Thr Gln Val
 355 360 365
 Ile Asn Pro Pro Asn Ser Ala Gln Lys Thr Glu Val Gln Pro Thr Gln
 370 375 380
 Val Ile Asp Gly Pro Phe Ala Gly Gly Lys Asp Thr Val Val Asn Ile
 385 390 395 400
 Asn Arg Ile Asn Thr Asn Ala Asp Gly Thr Ile Arg Val Gly Gly Phe
 405 410 415
 Lys Ala Ser Leu Thr Thr Asn Ala Ala His Leu His Ile Gly Lys Gly
 420 425 430
 Gly Val Asn Leu Ser Asn Gln Ala Ser Gly Arg Ser Leu Ile Val Glu
 435 440 445
 Asn Leu Thr Gly Asn Ile Thr Val Asp Gly Pro Leu Arg Val Asn Asn
 450 455 460
 Gln Val Gly Gly Tyr Ala Leu Ala Gly Ser Ser Ala Asn Phe Glu Phe
 465 470 475 480
 Lys Ala Gly Thr Asp Thr Lys Asn Gly Thr Ala Thr Phe Asn Asn Asp
 485 490 495

00360934 072600

Ile Ser Leu Gly Arg Phe Val Asn Leu Lys Val Asp Ala His Thr Ala
 500 505 510
 Asn Phe Lys Gly Ile Asp Thr Gly Asn Gly Gly Phe Asn Thr Leu Asp
 515 520 525
 Phe Ser Gly Val Thr Asp Lys Val Asn Ile Asn Lys Leu Ile Thr Ala
 530 535 540
 Ser Thr Asn Val Ala Val Lys Asn Phe Asn Ile Asn Glu Leu Ile Val
 545 550 555 560
 Lys Thr Asn Gly Ile Ser Val Gly Glu Tyr Thr His Phe Ser Glu Asp
 565 570 575
 Ile Gly Ser Gln Ser Arg Ile Asn Thr Val Arg Leu Glu Thr Gly Thr
 580 585 590
 Arg Ser Leu Phe Ser Gly Gly Val Lys Phe Lys Gly Gly Glu Lys Leu
 595 600 605
 Val Ile Asp Glu Phe Tyr Tyr Ser Pro Trp Asn Tyr Phe Asp Ala Arg
 610 615 620
 Asn Ile Lys Asn Val Glu Ile Thr Asn Lys Leu Ala Phe Gly Pro Gln
 625 630 635 640
 Gly Ser Pro Trp Gly Thr Ser Lys Leu Met Phe Asn Asn Leu Thr Leu
 645 650 655
 Gly Gln Asn Ala Val Met Asp Tyr Ser Gln Phe Ser Asn Leu Thr Ile
 660 665 670
 Gln Gly Asp Phe Ile Asn Asn Gln Gly Thr Ile Asn Tyr Leu Val Arg
 675 680 685
 Gly Gly Lys Val Ala Thr Leu Ser Val Gly Asn Ala Ala Ala Met Met
 690 695 700
 Phe Asn Asn Asp Ile Asp Ser Ala Thr Gly Phe Tyr Lys Pro Leu Ile
 705 710 715 720
 Lys Ile Asn Ser Ala Gln Asp Leu Ile Lys Asn Thr Glu His Val Leu
 725 730 735
 Leu Lys Ala Lys Ile Ile Gly Tyr Gly Asn Val Ser Thr Gly Thr Asn
 740 745 750
 Gly Ile Ser Asn Val Asn Leu Glu Glu Gln Phe Lys Glu Arg Leu Ala
 755 760 765
 Leu Tyr Asn Asn Asn Asn Arg Met Asp Thr Cys Val Val Arg Asn Thr
 770 775 780 785 790 795

00360934 072600
009220 42609260

770

775

780

Asp 785	Asp	Ile	Lys	Ala	Cys 790	Gly	Met	Ala	Ile	Gly 795	Asp	Gln	Ser	Met	Val 800
Asn	Asn	Pro	Asp	Asn 805	Tyr	Lys	Tyr	Leu	Ile 810	Gly	Lys	Ala	Trp	Lys 815	Asn
Ile	Gly	Ile	Ser 820	Lys	Thr	Ala	Asn	Gly 825	Ser	Lys	Ile	Ser	Val 830	Tyr	Tyr
Leu	Gly	Asn 835	Ser	Thr	Pro	Thr	Glu 840	Asn	Gly	Gly	Asn	Thr 845	Thr	Asn	Leu
Pro	Thr 850	Asn	Thr	Thr	Ser	Asn 855	Ala	Arg	Ser	Ala	Asn 860	Asn	Ala	Leu	Ala
Gln 865	Asn	Ala	Pro	Phe	Ala 870	Gln	Pro	Ser	Ala	Thr 875	Pro	Asn	Leu	Val	Ala 880
Ile	Asn	Gln	His	Asp 885	Phe	Gly	Thr	Ile	Glu 890	Ser	Val	Phe	Glu	Leu 895	Ala
Asn	Arg	Ser	Lys 900	Asp	Ile	Asp	Thr	Leu 905	Tyr	Ala	Asn	Ser	Gly 910	Ala	Gln
Gly	Arg 915	Asp	Leu	Leu	Gln	Thr	Leu 920	Leu	Ile	Asp	Ser	His 925	Asp	Ala	Gly
Tyr 930	Ala	Arg	Lys	Met	Ile	Asp 935	Ala	Thr	Ser	Ala	Asn 940	Glu	Ile	Thr	Lys
Gln 945	Leu	Asn	Thr	Ala	Thr 950	Thr	Thr	Leu	Asn	Asn 955	Ile	Ala	Ser	Leu	Glu 960
His	Lys	Thr	Ser	Gly 965	Leu	Gln	Thr	Leu	Ser 970	Leu	Ser	Asn	Ala	Met 975	Ile
Leu	Asn	Ser	Arg 980	Leu	Val	Asn	Leu	Ser 985	Arg	Arg	His	Thr	Asn 990	His	Ile
Asp	Ser 995	Phe	Ala	Lys	Arg	Leu	Gln 1000	Ala	Leu	Lys	Asp	Gln 1005	Lys	Phe	Ala
Ser 1010	Leu	Glu	Ser	Ala	Ala	Glu 1015	Val	Leu	Tyr	Gln	Phe 1020	Ala	Pro	Lys	Tyr
Glu 1025	Lys	Pro	Thr	Asn	Val 1030	Trp	Ala	Asn	Ala	Ile 1035	Gly	Gly	Thr	Ser	Leu 1040
Asn	Asn	Gly	Ser	Asn 1045	Ala	Ser	Leu	Tyr	Gly 1050	Thr	Ser	Ala	Gly	Val 1055	Asp

09360934.072699

Ala Tyr Leu Asn Gly Gln Val Glu Ala Ile Val Gly Gly Phe Gly Ser
1060 1065 1070

Tyr Gly Tyr Ser Ser Phe Asn Asn Arg Ala Asn Ser Leu Asn Ser Gly
1075 1080 1085

Ala Asn Asn Thr Asn Phe Gly Val Tyr Ser Arg Ile Phe Ala Asn Gln
1090 1095 1100

His Glu Phe Asp Phe Glu Ala Gln Gly Ala Leu Gly Ser Asp Gln Ser
1105 1110 1115 1120

Ser Leu Asn Phe Lys Ser Ala Leu Leu Gln Asp Leu Asn Gln Ser Tyr
1125 1130 1135

His Tyr Leu Ala Tyr Ser Ala Ala Thr Arg Ala Ser Tyr Gly Tyr Asp
1140 1145 1150

Phe Ala Phe Phe Arg Asn Ala Leu Val Leu Lys Pro Ser Val Gly Val
1155 1160 1165

Ser Tyr Asn His Leu Gly Ser Thr Asn Phe Lys Ser Asn Ser Thr Asn
1170 1175 1180

Gln Val Ala Leu Lys Asn Gly Ser Ser Ser Gln His Leu Phe Asn Ala
1185 1190 1195 1200

Ser Ala Asn Val Glu Ala Arg Tyr Tyr Tyr Gly Asp Thr Ser Tyr Phe
1205 1210 1215

Tyr Met Asn Ala Gly Val Leu Gln Glu Phe Ala His Val Gly Ser Asn
1220 1225 1230

Asn Ala Ala Ser Leu Asn Thr Phe Lys Val Asn Ala Ala Arg Asn Pro
1235 1240 1245

Leu Asn Thr His Ala Arg Val Met Met Gly Gly Glu Leu Lys Leu Ala
1250 1255 1260

Lys Glu Val Phe Leu Asn Leu Gly Val Val Tyr Leu His Asn Leu Ile
1265 1270 1275 1280

Ser Asn Ile Gly His Phe Ala Ser Asn Leu Gly Met Arg Tyr Ser Phe
1285 1290 1295

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5925 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

CTCCATTTTA AGCAACTCCA TAGACCACTA AAGAACTTT TTTTGAGGCT ATCTTTGAAA	60
ATCTGTCCTA TTGATTTGTT TTCCATTTTG TTTCCCATGT GGATCTTGTG GATCACAAAC	120
GCTTAATTAT ACATGCTATA GTAAGCATGA CACACAAACC AACTATTTTT TAGAACGCTT	180
CATGTGCTCA CCTTGACTAA CCATTTCTCC AACCATACTT TAGCGTTGCA TTTGATTTCT	240
TCAAAAAGAT TCATTTCTTA TTTCTTGTTT TTATTAAAGT TCTTTCATTT TAGCAAATTT	300
TTGTTAATTG TGGGTAAAAA TGTGAATCGT CCTAGCCTTT AGACGCCTGC AACGATCGGG	360
CTTTTTTCAA TATTAATAAT GATTAATGAA AAAAAAAAAA AATGCTTGAT ATTGTTGTAT	420
AATGAGAATG TTCAAAGACA TGAATTGACT ACTCAAGCGT GTAGCGATTT TTAGCAGTCT	480
TTGACACTAA CAAGATACCG ATAGGTATGA AACTAGGTAT AGTAAGGAGA AACAAATGACT	540
AACGAAACCA TTGACCAACA ACCACAAACC GAAGCGGCTT TTAACCCGCA GCAATTTATC	600
AATAATCTTC AAGTAGCTTT TCTTAAAGTT GATAACGCTG TCGCTTCATA CGATCCTGAT	660
CAAAAACCAA TCGTTGATAA GAACGATAGG GATAACAGGC AAGCTTTTGA AGGAATCTCG	720
CAATTAAGGG AAGAATACTC CAATAAAGCG ATCAAAAATC CTACCAAAAA GAATCAGTAT	780
TTTTCAGACT TTATCAATAA GAGCAATGAT TTAATCAACA AAGACAATCT CATTGATGTA	840
GAATCTTCCA CAAAGAGCTT TCAGAAATTT GGGGATCAGC GTTACCGAAT TTTCACAAGT	900
TGGGTGTCCC ATCAAAACGA TCCGTCTAAA ATCAACACCC GATCGATCCG AAATTTTATG	960
GAAAATATCA TACAACCCCC TATCCTTGAT GATAAAGAGA AAGCGGAGTT TTTGAAATCT	1020
GCCAAACAAT CTTTTGCAGG AATCATTATA GGGAATCAAA TCCGAACGGA TCAAAAGTTC	1080
ATGGGCGTGT TTGATGAGTC CTTGAAAGAA AGGCAAGAAG CAGAAAAAAA TGGAGAGCCT	1140
ACTGGTGGGG ATTGGTTGGA TATTTTTCTC TCATTTATAT TTGACAAAAA ACAATCTTCT	1200
GATGTCAAAG AAGCAATCAA TCAAGAACCA GTTCCCCATG TCCAACCAGA TATAGCCACT	1260
ACCACCACCG ACATACAAGG CTTACCGCCT GAAGCTAGAG ATTTACTTGA TGAAAGGGGT	1320

AATTTTCTA	AATTCACCT	TGGCGATATG	GAAATGTTAG	ATGTTGAGGG	AGTCGCTGAC	1380
ATTGATCCCA	ATTACAAGTT	CAATCAATTA	TTGATTCCACA	ATAACGCTCT	GTCTTCTGTG	1440
TTAATGGGGA	GTCATAATGG	CATAGAACCT	GAAAAAGTTT	CATTGTTGTA	TGGGGGCAAT	1500
GGTGGTCCTG	GAGCTAGGCA	TGATTGGAAC	GCCACCGTTG	GTTATAAAGA	CCAACAAGGC	1560
AACAATGTGG	CTACAATAAT	TAATGTGCAT	ATGAAAAACG	GCAGTGGCTT	AGTCATAGCA	1620
GGTGGTGAGA	AAGGGATTAA	CAACCCTAGT	TTTTATCTCT	ACAAAGAAGA	CCAACTCACA	1680
GGCTCACAAC	GAGCATTAA	TCAAGAAGAG	ATCCAAAACA	AAATAGATTT	CATGGAATTT	1740
CTTGACACAA	ATAATGCTAA	ATTAGACAAC	TTGAGCGAGA	AAGAGAAGGA	AAAATTCCGA	1800
ACTGAGATTA	AAGATTTCCA	AAAAGACTCT	AAGGCTTATT	TAGACGCCCT	AGGGAATGAT	1860
CGTATTGCTT	TTGTTTCTAA	AAAAGACACA	AAACATTTCAG	CTTTAATTAC	TGAGTTTGGT	1920
AATGGGGATT	TGAGCTACAC	TCTCAAAGAT	TATGGGAAAA	AAGCAGATAA	AGCTTTAGAT	1980
AGGGAGAAAA	ATGTTACTCT	TCAAGGTAGC	CTAAACATG	ATGGCGTGAT	GTTTGTTGAT	2040
TATTTCTAATT	TCAAATACAC	CAACGCCTCC	AAGAATCCCA	ATAAGGGTGT	AGGCGTTACG	2100
AATGGCGTTT	CCCATTTAGA	AGTAGGCTTT	AACAAGGTAG	CTATCTTTAA	TTTGCCTGAT	2160
TTAAATAATC	TCGCTATCAC	TAGTTTCGTA	AGGCGGAATT	TAGAGGATAA	ACTAACCCT	2220
AAAGGATTGT	CCCCACAAGA	AGCTAATAAG	CTTATCAAAG	ATTTTTTGAG	CAGCAACAAA	2280
GAAATGGTTG	GAAAAACTTT	AACTTCAAT	AAAGCTGTAG	CTGACGCTAA	AAACACAGGC	2340
AATTATGATG	AAGTGAAAAA	AGCTCAGAAA	GATCTTGAAA	AATCTCTAAG	GAAACGAGAG	2400
CATTTAGAGA	AAGAAGTAGA	GAAAAAATTG	GAGAGCAAAA	GCGGCAACAA	AAATAAAATG	2460
GAAGCAAAAG	CTCAAGCTAA	CAGCCAAAAA	GATGAGATTT	TTGCGTTGAT	CAATAAAGAG	2520
GCTAATAGAG	ACGCAAGAGC	AATCGCTTAC	GCTCAGAATC	TTAAAGGCAT	CAAAAGGGAA	2580
TTGTCTGATA	AACTTGAAAA	TGTCAACAAG	AATTTGAAAG	ACTTTGATAA	ATCTTTTGAT	2640
GAATTCAAAA	ATGGCAAAAA	TAAGGATTTT	AGCAAGGCAG	AAGAAACACT	AAAAGCCCTT	2700
AAAGGTTTCG	TGAAAGATTT	AGGTATCAAT	CCAGAATGGA	TTTCAAAAGT	TGAAAACCTT	2760
AATGCAGCTT	TGAATGAATT	CAAAAATGGC	AAAAATAAGG	ATTTCAAGCA	GGTAACGCAA	2820
GCAAAAAGCG	ACCTTGAAAA	TTCCGTAAAA	GATGTGATCA	TCAATCAAAA	GGTAACGGAT	2880

AAAGTTGATA	ATCTCAATCA	AGCGGTATCA	GTGGCTAAAG	CAACGGGTGA	TTTCAGTAGG	2940
GTAGAGCAAG	CGTTAGCCGA	TCTCAAAAAT	TTCTCAAAGG	AGCAATTGGC	CCAACAAGCT	3000
CAAAAAAATG	AAAGTCTCAA	TGCTAGAAAA	AAATCTGAAA	TATATCAATC	CGTTAAGAAT	3060
GGTGTGAATG	GAACCCTAGT	CGGTAATGGG	TTATCTCAAG	CAGAAGCCAC	AACTCTTTCT	3120
AAAAACTTTT	CGGACATCAA	GAAAGAGTTG	AATGCAAAAC	TTGGAAATTT	CAATAACAAT	3180
AACAATAATG	GACTCAAAAA	CGAACCCATT	TATGCTAAAG	TTAATAAAAA	GAAAGCAGGG	3240
CAAGCAGCTA	GCCTTGAAGA	ACCCATTTAC	GCTCAAGTTG	CTAAAAAGGT	AAATGCAAAA	3300
ATTGACCGAC	TCAATCAAAT	AGCAAGTGGT	TTGGGTGTTG	TAGGGCAAGC	AGCGGGCTTC	3360
CCTTTGAAAA	GGCATGATAA	AGTTGATGAT	CTCAGTAAGG	TAGGGCTTTC	AAGGAATCAA	3420
GAATTGGCTC	AGAAAATTGA	CAATCTCAAT	CAAGCGGTAT	CAGAAGCTAA	AGCAGGTTTT	3480
TTTGGCAATC	TAGAGCAAAC	GATAGACAAG	CTCAAAGATT	CTACAAAACA	CAATCCCATG	3540
AATCTATGGG	TTGAAAGTGC	AAAAAAAGTA	CCTGCTAGTT	TGTCAGCGAA	ACTAGACAAT	3600
TACGCTACTA	ACAGCCACAT	ACGCATTAAT	AGCAATATCA	AAAATGGAGC	AATCAATGAA	3660
AAAGCGACCG	GCATGCTAAC	GCAAAAAAAC	CCTGAGTGGC	TCAAGCTCGT	GAATGATAAG	3720
ATAGTTGCGC	ATAATGTAGG	AAGCGTTCCT	TTGTCAGAGT	ATGATAAAAT	TGGCTTCAAC	3780
CAGAAGAATA	TGAAAGATTA	TTCTGATTCG	TTCAAGTTTT	CCACCAAGTT	GAACAATGCT	3840
GTAAAAGACA	CTAATTCTGG	CTTTACGCAA	TTTTTAACCA	ATGCATTTTC	TACAGCATCT	3900
TATTACTGCT	TGGCGAGAGA	AAATGCGGAG	CATGGAATCA	AGAACGTAA	TACAAAAGGT	3960
GGTTTCCAAA	AATCTTAAAG	GATTAAGGAA	TACCAAAAAC	GCAAAAACCA	CCCCTTGCTA	4020
AAAGCGAGGG	GTTTTTTAAT	ACTCCTTAGC	AGAAATCCCA	ATCGTCTTTA	GTATTTGGGA	4080
TGAATGCTAC	CAATTCATGG	TATCATATCC	CCATACATTC	GTATCTAGCG	TAGGAAGTGT	4140
GCAAAGTTAC	GCCTTTGGAG	ATATGATGTG	TGAGACCTGT	AGGGAATGCG	TTGGAGCTCA	4200
AACTCTGTAA	AATCCCTATT	ATAGGGACAC	AGAGTGAGAA	CCAAACTCTC	CCTACGGGCA	4260
ACATCAGCCT	AGGAAGCCCA	ATCGTCTTTA	GCGGTTGGGC	ACTTCACCTT	AAAATATCCC	4320
GACAGACACT	AACGAAAGGC	TTTGTTCCTT	AAAGTCTGCA	TGGATATTTT	CTACCCCAAA	4380
AAGACTTAAC	CCTTTGCTTA	AAATTAAGTT	TGATTGTGCT	AGTGGGTTCG	TGCTATAGTG	4440

CGAAAATTAA	TTAAGGGTTA	TAAAGAGAGC	ATAAACTAGA	AAAAACAAGT	AGCTATAACA	4500
AAGATCAAGT	TCAAAAAATC	ATAGAGCTTT	TAGAGCAAAT	TGATCGCGCT	CTTAACCAAA	4560
GAAAAATCAG	AAAAACCATA	GGAATTATCA	CACCTTATAA	TGCCCAAAAA	AGACGCTTGC	4620
GATCAGAAGT	GGAAAAATAC	GGCTTCAAGA	ATTTTGATGA	GCTCAAATA	GACACTGTGG	4680
ATGCCTTTCA	AGGTGAAGAG	GCAGATATTA	TTATTTATT	CACCGTGAAA	ACTTGTGGTA	4740
ATCTTTCTTT	CTTGCTAGAT	TCTAAACGCT	TGAATGTGGC	TATTTCTAGG	GCAAAAGAAA	4800
ATCTCATTTT	TGTGGGTAAA	AAGTCTTTCT	TTGAGAATTT	ATGAAGCGAT	GAGAAGAATA	4860
TCTTTAGCGC	TATTTTGCAA	GTCTGTAGAT	AGGTAATCTT	TTCCAAAGAT	AATCATTAGA	4920
CATTCTTCGC	TTCAAACGC	TTTCATAAAT	CTCTCTAAAG	CGCTTTATAA	TCAACACAAT	4980
ACCTTATAG	TGTGAGCTAT	AGCCCCTTTT	TGGGAATTGA	GTTATTTTGA	CTTTAAATTT	5040
TTATTAGCGT	TACAATTTGA	GCCATTCTTT	AGCTTGTTTT	TCTAGCCAGA	TCACATCGCC	5100
GCTGCGCATGA	AATTCCACTT	TAGGGAATGC	GTGTGCATTT	TTTTTAAGGG	CGTATTTTTG	5160
CTGCAAATAT	CCTACAATAG	CATCGCCCGA	ATGGATGAGT	AGGGGGGGTG	TTGAAAGGGC	5220
AAAATGCTCC	ATAAAATAGC	CCTCAATTTT	TTGAGCGATT	AAGGGAAAAT	GCGTGCAACC	5280
TAAAATAATC	ACTTCGGGAA	AATCTTTAAG	GGAGTGAAAT	AATAACGCAT	GCAAGTTTCT	5340
AACAATTTCGC	CCTCTAAAAT	ACTTTCTTCA	ATCAAAGGCA	CAAAAAGAGA	AGTGGCTAAA	5400
TGCGAAACAT	TCAAATAGCC	TTGTTGTTTC	AGGGCATTGT	CATAAGCGTT	GGATTGGATC	5460
GTCGCTTTTG	TCCCTAGCAC	TAAAATAGGG	GCGTTTTTAT	CTTTTACTTG	TCGCTTGATC	5520
GCTAAAATGC	TTGGCTCAAT	CACGCCCA	ATAGGGATTT	TGGAATGCTT	TTGCATCTCT	5580
TCTAAAGCTA	GAGCGCTCGC	TGTGTTGCAT	GCCACAATCA	ATAATTCAAT	CTGGTGCGGT	5640
TTGAAAAAAT	CCAAAGCCTC	TAAGCCAAAT	TGCTTGATCG	TAGTGGGGTC	TTTAGTGCCA	5700
TAAGGCACTC	TAGCCGTATC	GCCATAATAG	ATGATTTTCAT	CAAATAATTG	CGCTTTTAAA	5760
AGGCTTTTTA	AAACGCTAAA	CCCTCCCACA	CCGCTATCAA	AAACGCCTAT	TTTCATGACA	5820
CTTTTTTAAT	TTAATGGGAT	TAATTAGGGA	TTTTATTTTT	CATTCATTAA	GTTTAAAAAT	5880
TCTTCATTGT	CCTTAGTTTG	TTGCATTTTA	GAATAGACAA	AGCTT		5925

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1147 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Met Thr Asn Glu Thr Ile Asp Gln Gln Pro Gln Thr Glu Ala Ala Phe
 1 5 10 15
 Asn Pro Gln Gln Phe Ile Asn Asn Leu Gln Val Ala Phe Leu Lys Val
 20 25 30
 Asp Asn Ala Val Ala Ser Tyr Asp Pro Asp Gln Lys Pro Ile Val Asp
 35 40 45
 Lys Asn Asp Arg Asp Asn Arg Gln Ala Phe Glu Gly Ile Ser Gln Leu
 50 55 60
 Arg Glu Glu Tyr Ser Asn Lys Ala Ile Lys Asn Pro Thr Lys Lys Asn
 65 70 75 80
 Gln Tyr Phe Ser Asp Phe Ile Asn Lys Ser Asn Asp Leu Ile Asn Lys
 85 90 95
 Asp Asn Leu Ile Asp Val Glu Ser Ser Thr Lys Ser Phe Gln Lys Phe
 100 105 110
 Gly Asp Gln Arg Tyr Arg Ile Phe Thr Ser Trp Val Ser His Gln Asn
 115 120 125
 Asp Pro Ser Lys Ile Asn Thr Arg Ser Ile Arg Asn Phe Met Glu Asn
 130 135 140
 Ile Ile Gln Pro Pro Ile Leu Asp Asp Lys Glu Lys Ala Glu Phe Leu
 145 150 155 160
 Lys Ser Ala Lys Gln Ser Phe Ala Gly Ile Ile Ile Gly Asn Gln Ile
 165 170 175
 Arg Thr Asp Gln Lys Phe Met Gly Val Phe Asp Glu Ser Leu Lys Glu
 180 185 190
 Arg Gln Glu Ala Glu Lys Asn Gly Glu Pro Thr Gly Gly Asp Trp Leu
 195 200 205
 Asp Ile Phe Leu Ser Phe Ile Phe Asp Lys Lys Gln Ser Ser Asp Val

00360934.072599

210					215					220					
Lys 225	Glu	Ala	Ile	Asn	Gln 230	Glu	Pro	Val	Pro	His 235	Val	Gln	Pro	Asp	Ile 240
Ala	Thr	Thr	Thr	Thr 245	Asp	Ile	Gln	Gly	Leu 250	Pro	Pro	Glu	Ala	Arg 255	Asp
Leu	Leu	Asp	Glu 260	Arg	Gly	Asn	Phe	Ser 265	Lys	Phe	Thr	Leu	Gly 270	Asp	Met
Glu	Met	Leu 275	Asp	Val	Glu	Gly	Val 280	Ala	Asp	Ile	Asp	Pro 285	Asn	Tyr	Lys
Phe 290	Asn	Gln	Leu	Leu	Ile	His 295	Asn	Asn	Ala	Leu	Ser 300	Ser	Val	Leu	Met
Gly 305	Ser	His	Asn	Gly	Ile 310	Glu	Pro	Glu	Lys	Val 315	Ser	Leu	Leu	Tyr	Gly 320
Gly	Asn	Gly	Gly	Pro 325	Gly	Ala	Arg	His	Asp 330	Trp	Asn	Ala	Thr	Val 335	Gly
Tyr	Lys	Asp	Gln 340	Gln	Gly	Asn	Asn	Val 345	Ala	Thr	Ile	Ile	Asn 350	Val	His
Met	Lys	Asn 355	Gly	Ser	Gly	Leu	Val 360	Ile	Ala	Gly	Gly	Glu 365	Lys	Gly	Ile
Asn 370	Asn	Pro	Ser	Phe	Tyr	Leu 375	Tyr	Lys	Glu	Asp	Gln 380	Leu	Thr	Gly	Ser
Gln 385	Arg	Ala	Leu	Ser	Gln 390	Glu	Glu	Ile	Gln	Asn 395	Lys	Ile	Asp	Phe	Met 400
Glu	Phe	Leu	Ala	Gln 405	Asn	Asn	Ala	Lys	Leu 410	Asp	Asn	Leu	Ser	Glu 415	Lys
Glu	Lys	Glu	Lys 420	Phe	Arg	Thr	Glu	Ile 425	Lys	Asp	Phe	Gln	Lys 430	Asp	Ser
Lys	Ala	Tyr 435	Leu	Asp	Ala	Leu	Gly 440	Asn	Asp	Arg	Ile	Ala 445	Phe	Val	Ser
Lys 450	Lys	Asp	Thr	Lys	His	Ser 455	Ala	Leu	Ile	Thr	Glu 460	Phe	Gly	Asn	Gly
Asp 465	Leu	Ser	Tyr	Thr	Leu 470	Lys	Asp	Tyr	Gly	Lys 475	Lys	Ala	Asp	Lys	Ala 480
Leu	Asp	Arg	Glu	Lys 485	Asn	Val	Thr	Leu	Gln 490	Gly	Ser	Leu	Lys	His 495	Asp

Gly Val Met Phe Val Asp Tyr Ser Asn Phe Lys Tyr Thr Asn Ala Ser
 500 505 510
 Lys Asn Pro Asn Lys Gly Val Gly Val Thr Asn Gly Val Ser His Leu
 515 520 525
 Glu Val Gly Phe Asn Lys Val Ala Ile Phe Asn Leu Pro Asp Leu Asn
 530 535 540
 Asn Leu Ala Ile Thr Ser Phe Val Arg Arg Asn Leu Glu Asp Lys Leu
 545 550 555 560
 Thr Thr Lys Gly Leu Ser Pro Gln Glu Ala Asn Lys Leu Ile Lys Asp
 565 570 575
 Phe Leu Ser Ser Asn Lys Glu Leu Val Gly Lys Thr Leu Asn Phe Asn
 580 585 590
 Lys Ala Val Ala Asp Ala Lys Asn Thr Gly Asn Tyr Asp Glu Val Lys
 595 600 605
 Lys Ala Gln Lys Asp Leu Glu Lys Ser Leu Arg Lys Arg Glu His Leu
 610 615 620
 Glu Lys Glu Val Glu Lys Lys Leu Glu Ser Lys Ser Gly Asn Lys Asn
 625 630 635 640
 Lys Met Glu Ala Lys Ala Gln Ala Asn Ser Gln Lys Asp Glu Ile Phe
 645 650 655
 Ala Leu Ile Asn Lys Glu Ala Asn Arg Asp Ala Arg Ala Ile Ala Tyr
 660 665 670
 Ala Gln Asn Leu Lys Gly Ile Lys Arg Glu Leu Ser Asp Lys Leu Glu
 675 680 685
 Asn Val Asn Lys Asn Leu Lys Asp Phe Asp Lys Ser Phe Asp Glu Phe
 690 695 700
 Lys Asn Gly Lys Asn Lys Asp Phe Ser Lys Ala Glu Glu Thr Leu Lys
 705 710 715 720
 Ala Leu Lys Gly Ser Val Lys Asp Leu Gly Ile Asn Pro Glu Trp Ile
 725 730 735
 Ser Lys Val Glu Asn Leu Asn Ala Ala Leu Asn Glu Phe Lys Asn Gly
 740 745 750
 Lys Asn Lys Asp Phe Ser Lys Val Thr Gln Ala Lys Ser Asp Leu Glu
 755 760 765
 Asn Ser Val Lys Asp Val Ile Ile Asn Gln Lys Val Thr Asp Lys Val

00360934.072699

770					775					780					
Asp 785	Asn	Leu	Asn	Gln	Ala 790	Val	Ser	Val	Ala	Lys 795	Ala	Thr	Gly	Asp	Phe 800
Ser	Arg	Val	Glu	Gln 805	Ala	Leu	Ala	Asp	Leu 810	Lys	Asn	Phe	Ser	Lys 815	Glu
Gln	Leu	Ala	Gln 820	Gln	Ala	Gln	Lys	Asn 825	Glu	Ser	Leu	Asn	Ala	Arg	Lys
Lys	Ser	Glu 835	Ile	Tyr	Gln	Ser	Val 840	Lys	Asn	Gly	Val	Asn 845	Gly	Thr	Leu
Val	Gly 850	Asn	Gly	Leu	Ser	Gln 855	Ala	Glu	Ala	Thr	Thr 860	Leu	Ser	Lys	Asn
Phe 865	Ser	Asp	Ile	Lys	Lys 870	Glu	Leu	Asn	Ala	Lys 875	Leu	Gly	Asn	Phe	Asn 880
Asn	Asn	Asn	Asn	Asn 885	Gly	Leu	Lys	Asn	Glu 890	Pro	Ile	Tyr	Ala	Lys 895	Val
Asn	Lys	Lys	Lys 900	Ala	Gly	Gln	Ala	Ala 905	Ser	Leu	Glu	Glu	Pro 910	Ile	Tyr
Ala	Gln	Val 915	Ala	Lys	Lys	Val	Asn 920	Ala	Lys	Ile	Asp	Arg 925	Leu	Asn	Gln
Ile 930	Ala	Ser	Gly	Leu	Gly	Val 935	Val	Gly	Gln	Ala	Ala 940	Gly	Phe	Pro	Leu
Lys 945	Arg	His	Asp	Lys	Val 950	Asp	Asp	Leu	Ser	Lys 955	Val	Gly	Leu	Ser	Arg 960
Asn	Gln	Glu	Leu	Ala 965	Gln	Lys	Ile	Asp	Asn 970	Leu	Asn	Gln	Ala	Val 975	Ser
Glu	Ala	Lys	Ala 980	Gly	Phe	Phe	Gly	Asn 985	Leu	Glu	Gln	Thr	Ile 990	Asp	Lys
Leu	Lys	Asp 995	Ser	Thr	Lys	His	Asn 1000	Pro	Met	Asn	Leu	Trp 1005	Val	Glu	Ser
Ala 1010	Lys	Lys	Val	Pro	Ala	Ser	Leu 1015	Ser	Ala	Lys	Leu	Asp	Asn	Tyr	Ala
Thr 1025	Asn	Ser	His	Ile	Arg 1030	Ile	Asn	Ser	Asn 1035	Ile	Lys	Asn	Gly	Ala	Ile 1040
Asn	Glu	Lys	Ala	Thr 1045	Gly	Met	Leu	Thr	Gln 1050	Lys	Asn	Pro	Glu	Trp 1055	Leu

00360934.072699

Lys Leu Val Asn Asp Lys Ile Val Ala His Asn Val Gly Ser Val Pro
 1060 1065 1070
 Leu Ser Glu Tyr Asp Lys Ile Gly Phe Asn Gln Lys Asn Met Lys Asp
 1075 1080 1085
 Tyr Ser Asp Ser Phe Lys Phe Ser Thr Lys Leu Asn Asn Ala Val Lys
 1090 1095 1100
 Asp Thr Asn Ser Gly Phe Thr Gln Phe Leu Thr Asn Ala Phe Ser Thr
 1105 1110 1115 1120
 Ala Ser Tyr Tyr Cys Leu Ala Arg Glu Asn Ala Glu His Gly Ile Lys
 1125 1130 1135
 Asn Val Asn Thr Lys Gly Gly Phe Gln Lys Ser
 1140 1145

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 546 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Met Ala Lys Glu Ile Lys Phe Ser Asp Ser Ala Arg Asn Leu Leu Phe
 1 5 10 15
 Glu Gly Val Arg Gln Leu His Asp Ala Val Lys Val Thr Met Gly Pro
 20 25 30
 Arg Gly Arg Asn Val Leu Ile Gln Lys Ser Tyr Gly Ala Pro Ser Ile
 35 40 45
 Thr Lys Asp Gly Val Ser Val Ala Lys Glu Ile Glu Leu Ser Cys Pro
 50 55 60
 Val Ala Asn Met Gly Ala Gln Leu Val Lys Glu Val Ala Ser Lys Thr
 65 70 75 80
 Ala Asp Ala Ala Gly Asp Gly Thr Thr Thr Ala Thr Val Leu Ala Tyr
 85 90 95
 Ser Ile Phe Lys Glu Gly Leu Arg Asn Ile Thr Ala Gly Ala Asn Pro
 100 105 110

Ile Glu Val Lys Arg Gly Met Asp Lys Ala Ala Glu Ala Ile Ile Asn
 115 120 125
 Glu Leu Lys Lys Ala Ser Lys Lys Val Gly Gly Lys Glu Glu Ile Thr
 130 135 140
 Gln Val Ala Thr Ile Ser Ala Asn Ser Asp His Asn Ile Gly Lys Leu
 145 150 155 160
 Ile Ala Asp Ala Met Glu Lys Val Gly Lys Asp Gly Val Ile Thr Val
 165 170 175
 Glu Glu Ala Lys Gly Ile Glu Asp Glu Leu Asp Val Val Glu Gly Met
 180 185 190
 Gln Phe Asp Arg Gly Tyr Leu Ser Pro Tyr Phe Val Thr Asn Ala Glu
 195 200 205
 Lys Met Thr Ala Gln Leu Asp Asn Ala Tyr Ile Leu Leu Thr Asp Lys
 210 215 220
 Lys Ile Ser Ser Met Lys Asp Ile Leu Pro Leu Leu Glu Lys Thr Met
 225 230 235 240
 Lys Glu Gly Lys Pro Leu Leu Ile Ile Ala Glu Asp Ile Glu Gly Glu
 245 250 255
 Ala Leu Thr Thr Leu Val Val Asn Lys Leu Arg Gly Val Leu Asn Ile
 260 265 270
 Ala Ala Val Lys Ala Pro Gly Phe Gly Asp Arg Arg Lys Glu Met Leu
 275 280 285
 Lys Asp Ile Ala Ile Leu Thr Gly Gly Gln Val Ile Ser Glu Glu Leu
 290 295 300
 Gly Leu Ser Leu Glu Asn Ala Glu Val Glu Phe Leu Gly Lys Ala Gly
 305 310 315 320
 Arg Ile Val Ile Asp Lys Asp Asn Thr Thr Ile Val Asp Gly Lys Gly
 325 330 335
 His Ser Asp Asp Val Lys Asp Arg Val Ala Gln Ile Lys Thr Gln Ile
 340 345 350
 Ala Ser Thr Thr Ser Asp Tyr Asp Lys Glu Lys Leu Gln Glu Arg Leu
 355 360 365
 Ala Lys Leu Ser Gly Gly Val Ala Val Ile Lys Val Gly Ala Ala Ser
 370 375 380
 Glu Val Glu Met Lys Glu Lys Lys Asp Arg Val Asp Asp Ala Leu Ser

09360934.072699

(2) INFORMATION FOR SEQ ID NO:7:

(A) LENGTH: 1838 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

AAGCTTGCTG	TCATGATCAC	AAAAAACACT	AAAAAACATT	ATTATTAAGG	ATACAAAATG	60
GCAAAAGAAA	TCAAATTTTC	AGATAGTGCG	AGAAACCTTT	TATTTGAAGG	CGTGAGGCAA	120
CTCCATGACG	CTGTCAAAGT	AACCATGGGG	CCAAGAGGCA	GGAATGTATT	GATCCAAAAA	180

AGCTATGGCG CTCCAAGCAT CACCAAAGAC GCGTGAGCG TGGCTAAAGA GATTGAATTA	240
AGTTGCCCAG TAGCTAACAT GGGCGCTCAA CTCGTTAAAG AAGTAGCGAG CAAAACCGCT	300
GATGCTGCCG GCGATGGCAC GACCACAGCG ACCGTGCTAG CTTATAGCAT TTTTAAAGAA	360
GGTTTGAGGA ATATCACGGC TGGGGCTAAC CCTATTGAAG TGAAACGAGG CATGGATAAA	420
GCTGCTGAAG CGATCATTAA TGAGCTTAAA AAAGCGAGCA AAAAAGTAGG CGGTAAAGAA	480
GAAATCACCC AAGTGGCGAC CATTCTGCA AACTCCGATC ACAATATCGG GAAACTCATC	540
GCTGACGCTA TGGAAAAAGT GGGTAAAGAC GCGTGATCA CCGTTGAGGA AGCTAAGGGC	600
ATTGAAGATG AATTGGATGT CGTAGAAGGC ATGCAATTG ATAGAGGCTA CCTCTCCCCT	660
TATTTTGTA CGAACGCTGA GAAAATGACC GCTCAATTGG ATAATGCTTA CATCCTTTTA	720
ACGGATAAAA AAATCTCTAG CATGAAAGAC ATTCTCCCGC TACTAGAAAA AACCATGAAA	780
GAGGGCAAAC CGCTTTTAAT CATCGCTGAA GACATTGAGG GCGAAGCTTT AACGACTCTA	840
GTGGTGAATA AATTAAGAGG CGTGTTGAAT ATCGCAGCGG TTAAAGCTCC AGGCTTTGGG	900
GACAGAAGAA AAGAAATGCT CAAAGACATC GCTATTTTAA CCGGCGGTCA AGTCATTAGC	960
GAAGAATTGG GCTTGAGTCT AGAAAACGCT GAAGTGGAGT TTTTAGGCAA AGCTGGAAGG	1020
ATTGTGATTG ACAAAGACAA CACCACGATC GTAGATGGCA AAGGCCATAG CGATGATGTT	1080
AAAGACAGAG TCGCGCAGAT CAAAACCCAA ATTGCAAGTA CGACAAGCGA TTATGACAAA	1140
GAAAAATTGC AAGAAAGATT GGCTAAACTC TCTGGCGGTG TGGCTGTGAT TAAAGTGGGC	1200
GCTGCGAGTG AAGTGGAAT GAAAGAGAAA AAAGACCGGG TGGATGACGC GTTGAGCGCG	1260
ACTAAAGCGG CGGTTGAAGA AGGCATTGTG ATTGGTGGCG GTGCGGCTCT CATTGCGCGG	1320
GCTCAAAAAG TGCATTTGAA TTTGCACGAT GATGAAAAAG TGGGCTATGA AATCATCATG	1380
CGCGCCATTA AAGCCCCATT AGCTCAAATC GCTATCAACG CTGGTTATGA TGGCGGTGTG	1440
GTCGTGAATG AAGTAGAAAA ACACGAAGGG CATTTTGGTT TTAACGCTAG CAATGGCAAG	1500
TATGTGGATA TGTTTAAAGA AGGCATTATT GACCCCTTAA AAGTAGAAAG GATCGCTCTA	1560
CAAATGCGG TTTCGGTTTC AAGCCTGCTT TTAACCACAG AAGCCACCGT GCATGAAATC	1620
AAAGAAGAAA AAGCGACTCC GGCAATGCCT GATATGGGTG GCATGGGCGG TATGGGAGGC	1680
ATGGGCGGCA TGATGTAAGC CCGCTTGCTT TTTAGTATAA TCTGCTTTTA AAATCCCTTC	1740

TCTAAATCCC CCCCTTTCTA AAATCTCTTT TTTGGGGGGG TGCTTTGATA AAACCGCTCG 1800
CTTGTA AAAA CATGCAACAA AAAATCTCTG TTAAGCTT 1838

09360934.072699